

Clustering Regression Wavelet Analysis for Lossless Compression of Hyperspectral Imagery

E. Ahanonu¹, M. W. Marcellin¹ and A. Bilgin^{1,2}

¹Dept. of Electrical & Computer Engineering, ²Dept. of Biomedical Engineering,
University of Arizona, Tucson, AZ 85721, USA

Abstract: Recently, Regression Wavelet Analysis (RWA) was proposed as a method for lossless compression of hyperspectral images. In RWA, a linear regression is performed after a spectral wavelet transform to generate predictors which estimate the detail coefficients from approximation coefficients at each scale of the spectral wavelet transform. In this work, we propose Clustering Regression Wavelet Analysis (RWA-C), an extension of the original ‘Restricted’ RWA model which may be used to improve compression performance while maintaining component scalability. We demonstrate that clustering may be used to group pixels with similar spectral profiles, these clusters may then be more efficiently processed to improve RWA prediction performance while only requiring a modest increase side-information.

1. Introduction

Remote sensing involves the acquisition of information for a spatial region from a distance. It is most commonly associated with the observation of Earth’s surface and is used in numerous applications ranging from ecology to military surveillance. Hyperspectral imaging is a type of remote sensing where information over a range of the electromagnetic spectrum is collected for each pixel. This information is then stored in the form of a multi-component image, where successive image components correspond to contiguous electromagnetic spectral bands. Depending on the properties of the sensor used to capture a hyperspectral image (HSI), there may be anywhere from a few hundred (e.g. 224 in AVIRIS [1]) to several thousand (e.g. 8461 in IASI [2]) components. This high component dimensionality results in hyperspectral imaging instruments producing massive amounts of data, which require efficient coding for practical storage and transmission.

Data compression techniques designed to reduce these resource requirements will generally be required to consider both spectral and spatial redundancies to achieve optimal bit-rate reductions. Due to its favorable energy compaction properties and high scalability, the Discrete Wavelet Transform (DWT) has been chosen as a decorrelating transform in several HSI coding frameworks [3] [4] [5]. Among these are the Part 2 extensions of JPEG2000 (JPEG2000-P2), which provide support for multi-component image transforms targeted at reducing cross-component dependencies. This results in significantly reduced bit-rate over JPEG2000 Part 1, where no cross-component decorrelation is defined for multi-spectral images.

1.1. Regression Wavelet Analysis (RWA)

To further improve the performance of DWT based HSI coding, Regression Wavelet Analysis (RWA) [6] was proposed to reduce remaining redundancies between wavelet coefficients after spectral wavelet transformation. This is achieved by performing a linear regression at each spectral DWT scale to generate a model which allows estimates of detail components to be generated using approximation components. RWA is divided into three models called *Maximum*, *Restricted*, and *Exogenous*. Each of these models is described below for an assumed HSI $\mathbf{X} \in \mathbb{R}^{m \times z}$, where m denotes the number of spatial pixels and z the number of spectral components. Additionally, RWA along with the proposed model are described for a single level of wavelet decomposition but are both easily generalizable for an arbitrary number of decomposition levels.

1.1.1 Maximum Model

In the *Maximum* model (*RWA-M*), all approximation components ($\mathbf{A}_1, \dots, \mathbf{A}_k$) at a given scale are used to generate an estimate ($\hat{\mathbf{D}}_i$) of the detail components (\mathbf{D}_i), this may be given by:

$$\hat{\mathbf{D}}_i = \beta_{i,0} + \sum_{n=1, \dots, k} \beta_{i,n} \mathbf{A}_n, \quad i = 1, \dots, k$$

where $\mathbf{D}_i, \hat{\mathbf{D}}_i, \mathbf{A}_n \in \mathbb{R}^{m \times 1}$, and $k = \frac{z}{2}$ is the number of subband components. While this model yields the best performance, the requirement of all approximation components results in poor scalability.

1.1.2 Restricted Model

The *Restricted* model (*RWA-R*) aims to preserve the component scalability of the DWT by using a reduced number of approximation components in the regression. The minimum number of required components will be determined by the filter used in the DWT. For the Haar filter, the RWA to preserve scalability will be given by:

$$\hat{\mathbf{D}}_i = \beta_{i,0} + \beta_{i,1} \mathbf{A}_i$$

In [6], they consider adding higher order terms to this regression to improve prediction, resulting in the following restricted model:

$$\hat{\mathbf{D}}_i = \beta_{i,0} + \beta_{i,1} \mathbf{A}_i + \beta_{i,2} (\mathbf{A}_i)^2 + \beta_{i,3} (\mathbf{A}_i)^3$$

Performance of the restricted model may be improved by adding additional approximation components into the regression, with the tradeoff being reduced scalability.

1.1.3 Exogenous Model

The *Exogenous (RWA-E)* model was introduced to relieve the computational burden of computing the linear least squares solution during encoding, as well as eliminating the need to send side information with each image. RWA-E works by training a linear model over multiple images, with the idea that images stemming from the same sensor will have similar spectral dependencies. The regression coefficients are then fixed and assumed to be consistent between the encoder and decoder.

2. Proposed Model

As mentioned above, *RWA-R* preserves the scalability of the DWT, while sacrificing compression performance due to the reduced available fitting data. To help minimize this loss, we propose the *Clustering Regression Wavelet Analysis (RWA-C)*. This model works by first dividing the pixels within an HSI into N clusters based on their spectral profiles. If the pixels in \mathbf{X} are given by $\mathbf{X} = [\mathbf{x}_1^T, \dots, \mathbf{x}_m^T]^T$ where $\mathbf{x}_i \in \mathbb{R}^{1 \times z}$, a cluster C_n is a partition of $M = \{1, 2, \dots, m\}$ such that the cardinality of C_n is m_n and $M = C_1 \cup \dots \cup C_N$. The approximation and detail components resulting from the spectral DWT of the pixels within a given cluster may be given by

$$\begin{aligned} S_{A_n} &= \{\mathbf{A}_1^n, \dots, \mathbf{A}_k^n\}, & \mathbf{A}_i^n &\in \mathbb{R}^{m_n \times 1} \\ &\text{and} \\ S_{D_n} &= \{\mathbf{D}_1^n, \dots, \mathbf{D}_k^n\}, & \mathbf{D}_i^n &\in \mathbb{R}^{m_n \times 1} \end{aligned}$$

respectively. For each set of approximation components S_{A_n} , an operation $F(\cdot)$ is applied to produce a single feature vector $\bar{\mathbf{a}}^n = F(S_{A_n}) \in \mathbb{R}^{1 \times k}$. This vector should represent the average profile across subband components for each pixel within cluster C_n , so that after applying $F(\cdot)$ to all S_{A_n} , we obtain a set of feature vectors $\bar{\mathbf{a}}^1, \dots, \bar{\mathbf{a}}^N$ which provide information about the average spectral profile for approximation components in each cluster. The *RWA-C* linear regression for an arbitrary detail component \mathbf{D}_i^n will be given by:

$$\hat{\mathbf{D}}_i^n = \beta_{i,0} + \beta_{i,1}\mathbf{A}_i^n + \beta_{i,2}(\mathbf{A}_i^n)^2 + \beta_{i,3}(\mathbf{A}_i^n)^3 + \sum_{q=1, \dots, k} \beta_{i,3+q} \bar{\mathbf{A}}_q^n$$

where $\bar{\mathbf{A}}_q^n \in \mathbb{R}^{m_n \times 1}$ has the form $\bar{\mathbf{A}}_q^n = [\bar{a}_q^n, \bar{a}_q^n, \dots, \bar{a}_q^n]^T$ where $\bar{a}_q^n \in \mathbb{R}$ is the q th element of the n th feature vector $\bar{\mathbf{a}}^n$. The coefficients β are then found through linear least square regression as in original *RWA*.

In summary, *RWA-C* classifies each pixel, and determines a feature vector which best describes the approximation subbands within each cluster. These feature vectors are used during linear regression to improve the estimate obtained using the restricted model, where the same feature vector is used for all pixels in the same class. The feature vectors are then included as side information along with the regression coefficients and pixel classifications. During decoding, a pixel of a given class uses the associated approximation components, along with its class feature vector to reproduce the detail component estimate.

2.1. Side Information

A DWT level consisting of k -components will require $(k + 4)k$ regression coefficients and N k -dimensional feature vectors. If we consider an AVIRIS image with $m = 500 * 500$ pixels, the side information associated with the first DWT level will only be 0.0015 bpppc when using 512 clusters (assuming single-precision regression coefficients). The largest side information is associated with the class labeling of each pixel, which will require $m \cdot \log_2(N)$ bits. For the same image as above and also assuming 512 cluster, the cost is roughly 0.04 bpppc.

3. Implementation

The validity of the proposed model is demonstrated by implementing the predictor and performing an analysis of the first-order entropy of the prediction residuals.

3.1. Clustering

To assign a class to each pixel we choose the k-means clustering algorithm using a squared Euclidean distance metric evaluated over 10 iterations. We provide results for 16, 32, 64, 128, 256 and 512 clusters. On average the time spent for clustering in each of these cases was 12, 19, 30, 53, 97, 189 seconds, respectively.

3.2. Feature Vector Extraction

While more sophisticated techniques may be used, we find that choosing the feature vectors to be the average spectral profile of approximation subbands within a cluster to provide enough information to achieve appreciable entropy reductions.

4. Experimental Results

Figure 1 shows the entropy of the prediction residuals for implementations using cluster sizes 16, 32, 64, 128, 256 and 512. As the number of k-mean clusters increases, the feature vector for each cluster better represents the characteristics of the approximation subbands, resulting in better prediction and reduced residual entropy. As observed in [6], performing spatial transformation after spectral decorrelation may result in an increased bit-rate. Our results show that spatial transformation results in an increase in entropy in all cases for RWA-M, but is image dependent for RWA-R and RWA-C.

Table 1 and 2 show the results over the test image set for DWT levels 1 and 2, respectively. In both figures, the top row of values within each cell represents entropy after spectral transform, the bottom row represents entropy after both spectral and spatial transform. In both Figure 1 as well as Table 1-2 the entropy for each subband considers the side information required for decodability. When using only a spectral transformation, the average reduction seen by RWA-C over RWA-R is 5.3% and 8.4% for DWT levels 1 and 2, respectively; where as if a spatial transform is applied after spectral transform the average is 0.7% and 1.1%, respectively. If the optimal transform technique is chosen for each method (Spectral-Only or Spectral-Spatial) the reductions are 2.4% and 1.7%, respectively. For calibrated images these reductions are 3.5% and 3.7%, while for uncalibrated images they are 1.6% and 0.2%. This is to be expected since the spectral profiles of uncalibrated pixels can be erratic, resulting in poor k-means convergence and increased spectral variance within each cluster.

In Figure 2 an example is provided showing the clusters resulting from performing the k-means algorithm on Yellowstone 11 (Calibrated) using 16 centers. The gray lines represent the approximation component profiles while the red lines give the resulting feature vector when using the mean as a feature vector extraction method. This figure shows that in some cases the feature vector well represents the average profile of all components, where as in other cases the profile variance among components within the cluster results in a feature vector which is a weak representation. We also observe in this example that the choice of N for the k-means algorithm should be chosen careful as to avoid ‘cluster overfitting’. That is, when looking at Figure 2 it is clear there are clusters whose members follow very similar spectral profiles. This will result in excess side information being sent in the form of feature vectors and bit require to represent class labels. A refined compression framework should incorporate a method to avoid this overfitting to avoid unnecessary computation and side-information.

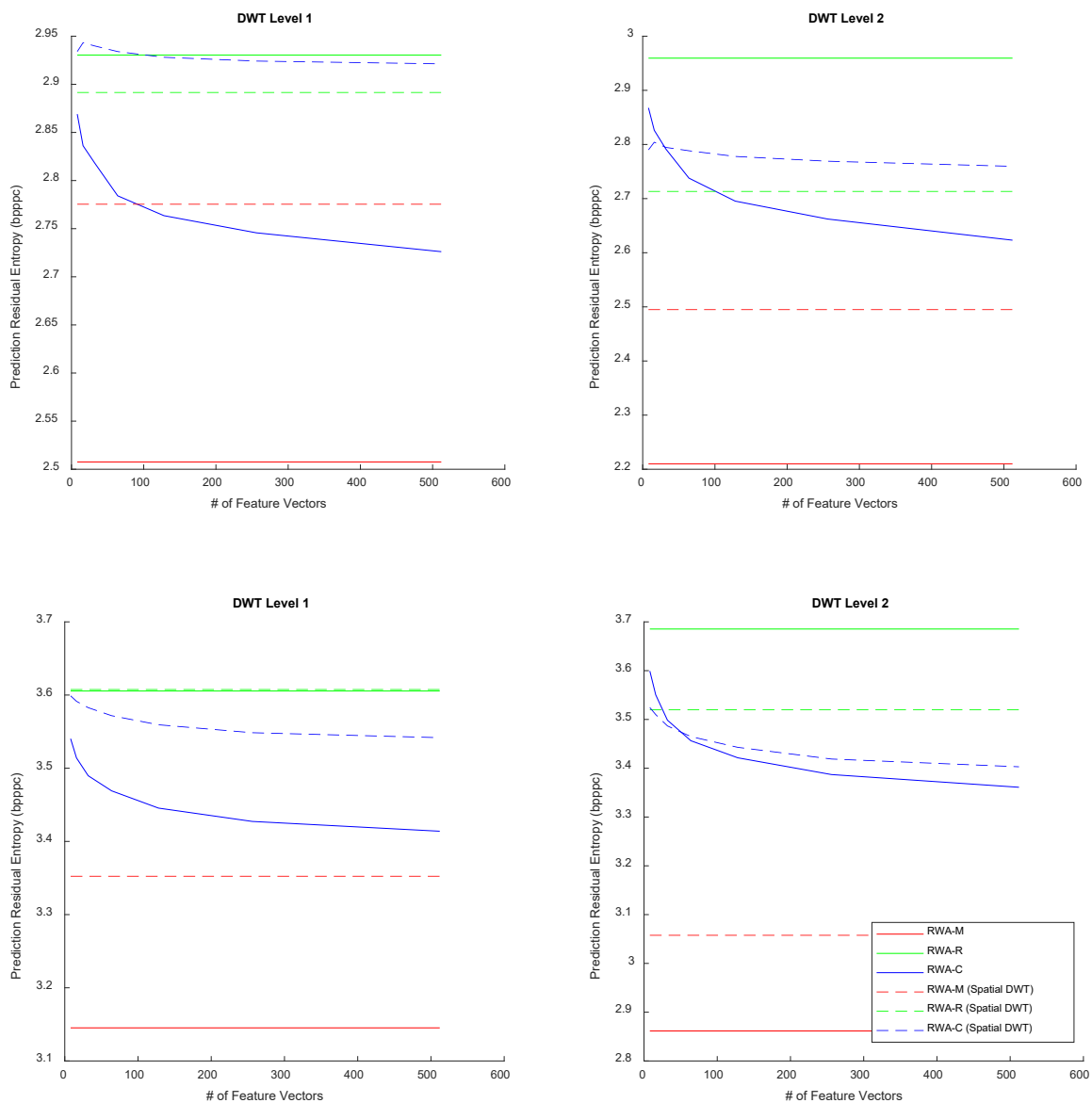


Figure 1. Entropy of prediction residuals over a range of k-means cluster counts, for Hawaii 1 (top) and Yellowstone 10 (bottom)

Table 1 Entropy of prediction residuals on DWT level 1.

Image		Entropy of Prediction Residuals (bpppc)							
		RWA-M	RWA-R	RWA-C					
				Cluster Size					
				16	32	64	128	256	512
Yellowstone 0 (Calibrated)	w/o spatial	3.52	4.39	4.24	4.20	4.16	4.13	4.11	4.08
	w/spatial	3.72	4.13	4.16	4.14	4.13	4.11	4.10	4.08
Yellowstone 0 (Uncalibrated)	w/o spatial	5.72	6.56	6.47	6.44	6.42	6.39	6.37	6.34
	w/spatial	5.89	6.27	6.33	6.32	6.33	6.32	6.32	6.31
Yellowstone 3 (Calibrated)	w/o spatial	3.41	4.16	4.03	3.99	3.96	3.92	3.90	3.87
	w/spatial	3.61	4.02	4.01	3.99	3.98	3.96	3.95	3.93
Yellowstone 3 (Uncalibrated)	w/o spatial	5.60	6.32	6.25	6.23	6.20	6.18	6.16	6.14
	w/spatial	5.77	6.16	6.19	6.18	6.18	6.17	6.17	6.15
Yellowstone 10 (Calibrated)	w/o spatial	3.15	3.61	3.52	3.49	3.47	3.44	3.43	3.41
	w/spatial	3.35	3.61	3.59	3.58	3.58	3.56	3.55	3.54
Yellowstone 10 (Uncalibrated)	w/o spatial	5.35	5.78	5.74	5.72	5.70	5.69	5.68	5.66
	w/spatial	5.51	5.75	5.76	5.76	5.76	5.75	5.75	5.74
Yellowstone 11 (Calibrated)	w/o spatial	3.32	3.95	3.81	3.75	3.72	3.70	3.69	3.66
	w/spatial	3.52	3.81	3.81	3.78	3.78	3.76	3.76	3.75
Yellowstone 11 (Uncalibrated)	w/o spatial	5.51	6.11	6.01	5.97	5.95	5.93	5.91	5.89
	w/spatial	5.68	5.95	5.97	5.96	5.96	5.95	5.95	5.94
Yellowstone 18 (Calibrated)	w/o spatial	3.41	4.38	4.18	4.12	4.09	4.04	4.01	3.98
	w/spatial	3.61	4.11	4.09	4.08	4.07	4.04	4.03	4.01
Yellowstone 18 (Uncalibrated)	w/o spatial	5.60	6.57	6.40	6.37	6.34	6.31	6.28	6.25
	w/spatial	5.78	6.27	6.28	6.27	6.27	6.25	6.25	6.23
Hawaii 1 (Uncalibrated)	w/o spatial	2.51	2.93	2.84	2.82	2.79	2.76	2.75	2.73
	w/spatial	2.78	2.89	2.95	2.94	2.94	2.93	2.93	2.92
Maine 10 (Uncalibrated)	w/o spatial	2.57	3.09	2.97	2.94	2.93	2.91	2.90	2.88
	w/spatial	2.82	2.97	3.00	3.00	3.00	3.00	3.00	3.00
AVERAGE	w/o spatial	4.14	4.82	4.69	4.65	4.61	4.59	4.56	4.53
	w/spatial	4.34	4.66	4.66	4.65	4.64	4.62	4.61	4.59

Table 2 Entropy of prediction residuals for DWT level 2

Image		Entropy of Prediction Residuals (bpppc)							
		RWA-M	RWA-R	RWA-C					
				Cluster Size					
				16	32	64	128	256	512
Yellowstone 0 (Calibrated)	w/o spatial	3.28	4.73	4.52	4.45	4.40	4.34	4.30	4.25
	w/spatial	3.44	4.19	4.23	4.20	4.18	4.15	4.13	4.10
Yellowstone 0 (Uncalibrated)	w/o spatial	5.44	6.88	6.77	6.72	6.68	6.63	6.60	6.56
	w/spatial	5.56	6.28	6.39	6.38	6.39	6.38	6.37	6.35
Yellowstone 3 (Calibrated)	w/o spatial	3.17	4.47	4.27	4.19	4.14	4.08	4.03	3.98
	w/spatial	3.33	4.08	4.08	4.04	4.01	3.98	3.95	3.91
Yellowstone 3 (Uncalibrated)	w/o spatial	5.32	6.59	6.50	6.46	6.43	6.38	6.35	6.31
	w/spatial	5.43	6.15	6.22	6.22	6.22	6.20	6.19	6.17
Yellowstone 10 (Calibrated)	w/o spatial	2.86	3.69	3.55	3.50	3.46	3.42	3.39	3.36
	w/spatial	3.06	3.52	3.51	3.48	3.47	3.44	3.42	3.40
Yellowstone 10 (Uncalibrated)	w/o spatial	4.99	5.81	5.74	5.71	5.69	5.66	5.64	5.61
	w/spatial	5.12	5.58	5.61	5.61	5.61	5.60	5.59	5.57
Yellowstone 11 (Calibrated)	w/o spatial	3.04	4.15	3.91	3.84	3.79	3.75	3.72	3.68
	w/spatial	3.21	3.76	3.76	3.73	3.70	3.68	3.66	3.64
Yellowstone 11 (Uncalibrated)	w/o spatial	5.17	6.28	6.14	6.08	6.04	6.00	5.97	5.93
	w/spatial	5.29	5.84	5.88	5.87	5.87	5.85	5.85	5.83
Yellowstone 18 (Calibrated)	w/o spatial	3.16	4.74	4.45	4.38	4.31	4.24	4.19	4.15
	w/spatial	3.33	4.23	4.22	4.20	4.16	4.12	4.09	4.05
Yellowstone 18 (Uncalibrated)	w/o spatial	5.32	6.91	6.69	6.65	6.60	6.56	6.52	6.47
	w/spatial	5.44	6.36	6.40	6.38	6.37	6.35	6.34	6.31
Hawaii 1 (Uncalibrated)	w/o spatial	2.21	2.96	2.83	2.79	2.74	2.69	2.67	2.62
	w/spatial	2.49	2.71	2.81	2.79	2.79	2.78	2.77	2.76
Maine 10 (Uncalibrated)	w/o spatial	2.34	3.24	3.03	2.99	2.96	2.93	2.91	2.88
	w/spatial	2.57	2.83	2.89	2.88	2.89	2.88	2.88	2.87
AVERAGE	w/o spatial	3.86	5.04	4.85	4.79	4.74	4.69	4.65	4.61
	w/spatial	4.02	4.63	4.65	4.63	4.60	4.59	4.56	4.54

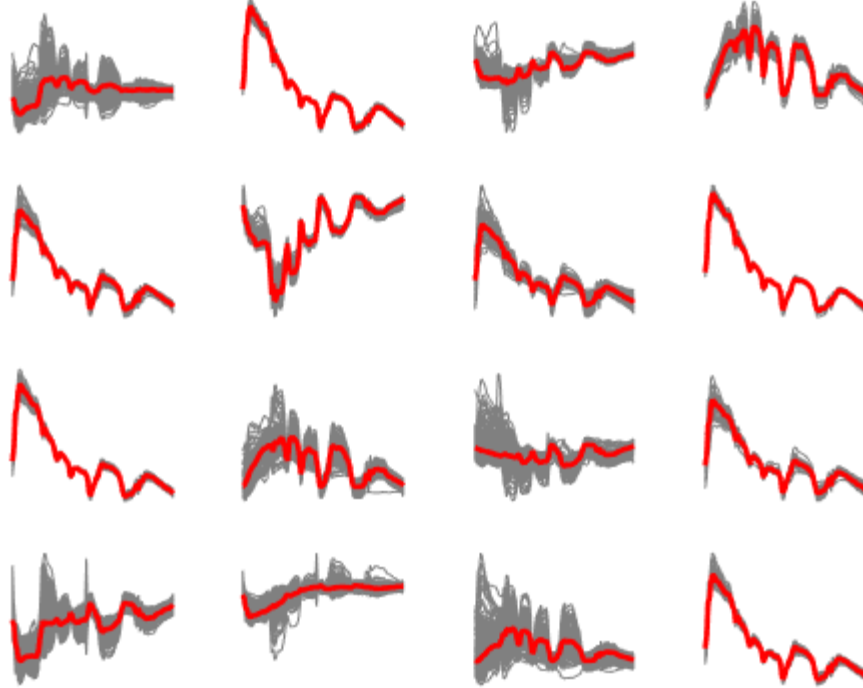


Figure 2 Examples of approximation component clusters (gray) and resulting feature vectors (red) on Yellowstone 11 (Calibrated) using K-mean clustering with 16 centers

5. Conclusions

In this paper, we proposed the *Clustering Regression Wavelet Analysis (RWA-C)* model which looks to group pixel with similar spectral profiles into clusters. A feature vector is extracted for each cluster to describe the general characteristics of the approximation subbands within each cluster. Rather than using all approximation components for each pixel in linear regression, we instead used the feature vector for each pixel, along with the approximation components required to preserve scalability of the DWT. We demonstrated that this model may be used to reduce bit-rate requirements in applications in which component scalability is preferred.

5.1. Future Work

This model relies on the feature vector for each cluster containing sufficient information to improve linear regression of each component. Here, we demonstrated that using the average spectral profile of approximation components in each cluster may yield positive entropy reductions. However, it is likely that a more sophisticated technique may be used to produce this feature vector, such as an artificial neural network (ANN). Additionally, the linear regression may be replaced with a non-linear method which can better exploit the information within the feature vector. Ideally, there would be two ANNs where one would produce an optimal feature vector that the other could use to perform regression.

References

- [1] J. P. Laboratory, "Airborne Visible InfraRed Imaging Spectrometer," [Online]. Available: <https://aviris.jpl.nasa.gov>.
- [2] "Infrared atmospheric sounding interferometer (IASI)," [Online]. Available: <https://wdc.dlr.de/sensors/iasi>.
- [3] Consultative Committee for Space Data Systems (CCSDS), "Lossless Multispectral & Hyperspectral Image Compression CCSDS 123.0-B-1, ser. Blue Book," May 2012. [Online]. Available: <https://public.ccsds.org/Pubs/123x0b1ec1.pdf>.
- [4] X. Tang, W. Pearlman and J. Modestino, "Hyperspectral image compression using three-dimensional wavelet coding," in *Proceedings of SPIE*, 2003.
- [5] B. Töreyn, O. Yilmaz, Y. Mert and F. Türk, "Lossless hyperspectral image compression using wavelet transform based spectral decorrelation," in *7th International Conference on Recent Advances in Space Technologies (RAST)*, Istanbul, 2015.
- [6] N. Amrani, J. Serra-Sagristà, V. Laparra, M. Marcellin and J. Malo, "Regression Wavelet Analysis for Lossless Coding of Remote-Sensing Data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 9, pp. 5616-5627, 2016.